

AD-A169 972

EXTREMAL PROCESSES RECORD TIMES AND STRONG
APPROXIMATION(U) NORTH CAROLINA UNIV AT CHAPEL HILL
CENTER FOR STOCHASTIC PROCESSES D PFEIFER DEC 85

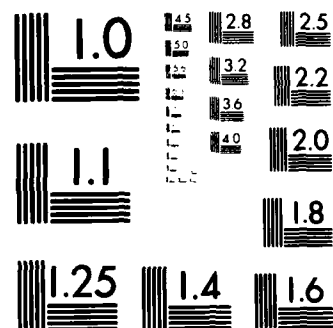
1/1

UNCLASSIFIED

TR-131 AFOSR-TR-86-0340 F49620-85-C-0144 F/G 12/1

NL



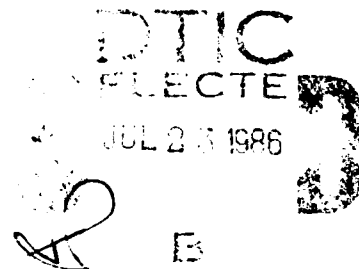
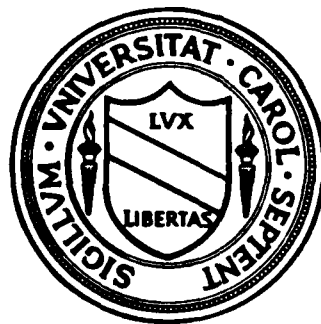


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

CENTER FOR STOCHASTIC PROCESSES

AD-A169 972

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina



EXTREMAL PROCESSES, RECORD TIMES AND STRONG APPROXIMATION

by

Dietmar Pfiefer

Technical Report No. 131

December 1985

DTIC FILE COPY

Approved for public release;
distribution is unlimited.

ALL INFORMATION CONTAINED HEREIN IS UNCLASSIFIED

DECLASSIFICATION/DOWNGRADING SCHEDULE

AFOSR-TR-86-0340
(FORMING ORGANIZATION REPORT NUMBER)

2. NAME OF MONITORING ORGANIZATION
AFOSR/NM

b. ADDRESS (City, State and ZIP Code)

PHOCUMENT IDENTIFICATION NUMBER
E49620 85 C 0144

PROGRAM ELEMENT NO.	6.1102F	TOXIMATION"
PROJECT NO.	2304	
TASK NO.	AS	
WORK UNIT NO.		

18. SUPPLEMENTARY NOTATION

13b. TIME COVERED

18. PAGE COUNT	13
19. DATE OF REPORT (Yr., Mo., Day)	December 1985

FIELD	GROUP	SUB. GR.
-------	-------	----------

SYNOPSIS: Extremal process, record times, strong approximation, Poisson approximation.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Given an i.i.d. sequence of random variables (r.v.'s) with continuous cumulative distribution function (CDF) F , we present a simple construction for the jump times of an extremal process on the same probability space which 'interpolate' the given record times. This gives another approach to the strong approximation of extremal processes as developed by Deheuvels (1981, 1982, 1983), and allows for a more detailed investigation of the relationship between the record times of the given sequence and the jump times of the extremal process. In particular, it is shown that the number S of surplus jump time points in $(1, \infty)$ over the record times is approximately Poisson distributed with an exact mean of $E(S) = 1 - C$, C denoting Euler's constant.

NO DISTRIBUTION/AVAILABILITY OF ABSTRACT

☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

UNCLASSIFIED

22b. TELEPHONE NUMBER (200)
(Include Area Code) 919-962-2907
22c. OFFICE SYMBOL AFOSR/NM

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

DD FORM 1473, 83 APR

EDITION OF 1 JAN 73 IS OBSOLETE.

98

9.

②

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
NOTICE OF TRANSMISSION TO DTIC
This technical report is approved for release and is
approved for release under E.O. 12958-12.
Distribution Statement
MATTHEW J. KLEIN
Chief, Technical Information Division

EXTREMAL PROCESSES, RECORD TIMES AND STRONG APPROXIMATION

by
Dietmar Pfeifer

Technical University Aachen
and

University of North Carolina at Chapel Hill

DTIC
ELECTE
JUL 23 1986
B

Abstract

Given an i.i.d. sequence of random variables (r.v.'s) with continuous cumulative distribution function (CDF) F , we present a simple construction for the jump times of an extremal process on the same probability space which 'interpolate' the given record times. This gives another approach to the strong approximation of extremal processes as developed by Deheuvels (1981, 1982, 1983), and allows for a more detailed investigation of the relationship between the record times of the given sequence and the jump times of the extremal process. In particular, it is shown that the number S of surplus jump time points in $(1, \infty)$ over the record times is approximately Poisson distributed with an exact mean of $E(S) = 1 - C$, C denoting Euler's constant.

Keywords: Extremal process, record times, strong approximation, Poisson approximation.

This research supported by Air Force Office of Scientific Research Grant F49620 85C 0144.

1. Introduction

Let $\{X_n; n \geq 1\}$ be an i.i.d. sequence of r.v.'s with continuous CDF F , and let $X_{(n)} = \max\{X_1, \dots, X_n\}$, $n \geq 1$. Of particular interest are the times U_n when these partial maxima change their values, defined by

$$U_0 = 1, U_{n+1} = \inf\{k \mid X_k > X_{U_n}\}, n \geq 0. \quad (1.1)$$

Due to the continuity of F , (1.1) is a.s. well-defined; U_n is also called the n^{th} record time, and X_{U_n} the n^{th} record value of the sequence. Several efforts have been made to clarify the asymptotic properties of record times, using different approaches such as canonical representations ([12],[13]), strong approximation techniques ([3],[6]) or embedding into extremal processes ([7],[9]), all of them saying that $\{\log U_n; n \geq 1\}$ asymptotically behaves like a homogeneous Poisson point process with unit rate.

Here the extremal process $\{E(t); t > 0\}$ (called extremal - F) is a right continuous non-decreasing pure jump Markov process such that for all selections $0 < t_1 < \dots < t_k$ of time points we have

$$P\left(\bigcap_{i=1}^k \{E(t_i) \leq x_i\}\right) = F^{t_1}(\min\{x_1, \dots, x_n\}) \prod_{i=2}^k F^{t_k - t_{k-1}}(\min\{x_i, \dots, x_k\}) \quad (1.2)$$

where $x_1, \dots, x_k \in \mathbb{R}$. Especially, from (1.2) it follows that we have

$\{X_{(1)}, \dots, X_{(n)}\} \stackrel{L}{=} \{E(1), \dots, E(n)\}$ for all $n \geq 1$, where $\stackrel{L}{=}$ means equality in distribution. The structural properties of such extremal processes are well-investigated (cf. [7]-[10]), and their importance is given by the fact that they occur as the functional weak limits of the normalized processes

$\left\{\frac{1}{b_n}(X_{([nt])} - a_n); t > 0\right\}$ ($[\cdot]$ denoting integer part) where $a_n \in \mathbb{R}$, $b_n > 0$ are

constants such that $\frac{1}{b_n}(X_{(n)} - a_n)$ tends weakly to an extreme value distribution (if applicable) (see e.g. [9] and further references therein). Further, if $\{\tau_n; -\infty < n < \infty\}$ denotes the jump times of the extremal-F process, it has been shown that these form a non-homogeneous Poisson point process with intensity $\lambda(t) = 1/t$, $t > 0$ (in fact, the extremal process has infinitely many jumps in every neighborhood of the origin). Correspondingly, the sequence $\{E(\tau_n); -\infty < n < \infty\}$ of states visited forms a Markov chain with transition probabilities

$$P(E(\tau_{n+1}) > y | E(\tau_n) = x) = \frac{-\log F(y)}{-\log F(x)}, \quad y \geq x \quad (1.3)$$

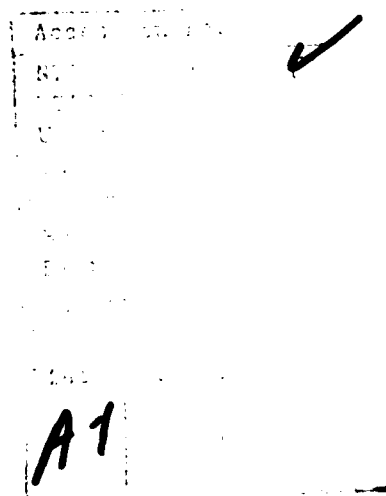
where x, y are chosen such that $0 < F(x) \leq F(y) < 1$. Since the distribution of $\{\tau_n\}$ is independent of F , $\{E(t)\}$ can be transformed to an extremal- Λ process $\{E^*(t)\}$ by letting $E^*(\tau_n) = -\log\{-\log F(E(\tau_n))\}$, $t > 0$, and interpolating with piecewise constant paths, where $\Lambda(x) = \exp(e^{-x})$, $x \in \mathbb{R}$ is the CDF of a doubly exponential distribution. Then $\{E^*(\tau_n)\}$ forms a homogeneous Poisson process on \mathbb{R} with unit rate. It follows that the time-transformed process $\{E^*(e^t); t \in \mathbb{R}\}$ now is homogeneous Poisson both in time and space.

In the light of (1.2), one might ask whether extremal processes can also be constructed by some sort of extension of the partial maxima (or records) from the original sequence, on the same space. Such considerations have been recently made by Deheuvels ([1],[2]) who started with a strong approximation of the record times $\{U_n\}$, which he then extended to a strong approximation of the inverse extremal process, and finally to the extremal process itself. We shall show in this paper that also a 'direct' approach is possible, constructing first the extremal jump times from the given record times.

This also allows for a more detailed investigation of the relationship between the jump times $\{\tau_n\}$ of the extremal process and the record times $\{U_n\}$, completing the results of Resnick ([7],[9]). In particular, if S denotes the number of surplus points in the τ_k -sequence over the record times, counted in the interval $(1, \infty)$, then S is approximately Poisson distributed, with mean

$$E(S) = 1 - C, \quad (1.4)$$

where $C = .577$ denotes Euler's constant. An estimation for the distance of S and the approximating Poisson r.v. in terms of total variation is also given.



2. Construction of the extremal jumps

In view of what has been said earlier, it is easier to work with the time-transformed process $\{E(e^t); t \in \mathbb{R}\}$ since then the corresponding jump times $\{\log \tau_n; -\infty < n < \infty\}$ form a homogeneous Poisson process with unit rate. Further, by the general structure of extremal processes, the jump times $\{\log \tau_n\}$ must be a.s. concentrated in the random set $\bigcup_{k=1}^{\infty} (\log(U_k - 1), \log U_k)$. In fact, in our construction, $\log \tau_1 \in (\log(U_1 - 1), \log U_1)$.

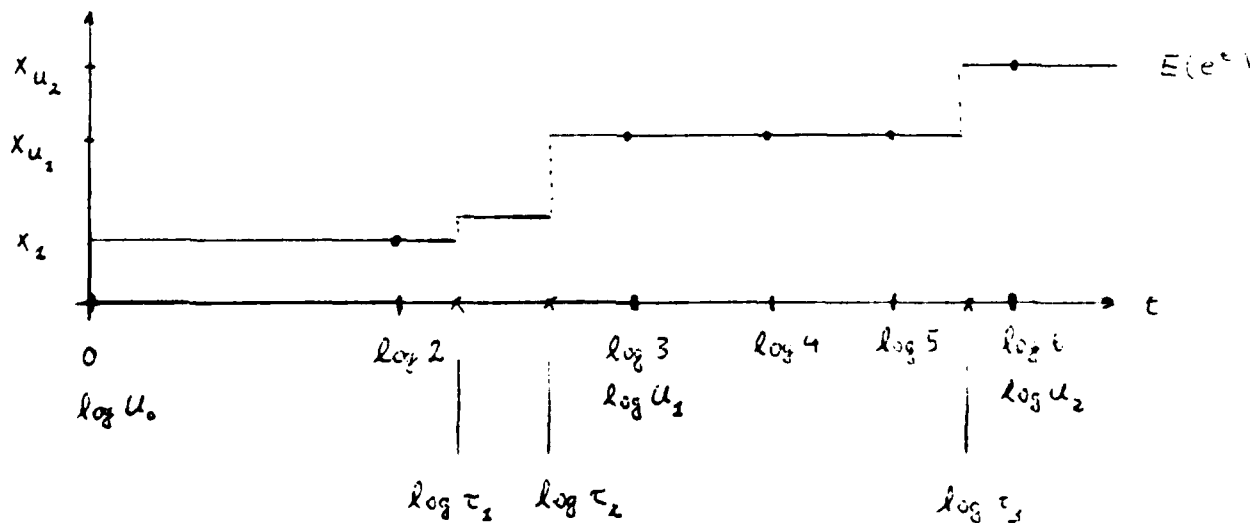
Let, for real numbers $a < b$, $N(a, b)$ denote the number of $(\log \tau)$ -points in the interval (a, b) . As a simple consequence of the Poissonian nature of $\{\log \tau_n; n \geq 1\}$, in a successful construction, the random variables $A_k = N(\log(U_k - 1), \log U_k)$ should be conditionally independent given the σ -field $A = A(U_1, U_2, \dots)$, following a (below) truncated Poisson distribution $Q(\lambda_k)$, say, with parameter

$$\lambda_k = \log\left(\frac{U_k}{U_k - 1}\right) \quad (2.1)$$

where

$$Q(\lambda, j) = \frac{1}{e^\lambda - 1} \frac{\lambda^j}{j!}, \quad j = 1, 2, 3, \dots, (\lambda > 0). \quad (2.2)$$

Further, conditioned on A and the number A_k , the location of the points in $(\log(U_k - 1), \log U_k)$ should be distributionally the same as that of an ordered sample of a population distributed uniformly over this interval.



By means of two independent i.i.d. uniformly $U[0,1]$ -distributed sequences $\{W_n(i); n \in \mathbb{N}\}$, $i = 1, 2$ (which can independently of $\{X_n\}$ be defined on the same probability space, eventually after enlarging) we are thus able to interpolate the given record times by extremal jumps in the following way.

Step 1. Determination of number of jumps.

Let $F_{Q(\lambda)}$ denote the CDF of the truncated Poisson distribution with parameter $\lambda > 0$. Define

$$A_k = F_{Q(\lambda_k)}^{-1}(W_k(1)), \quad k \in \mathbb{N}, \quad (2.3)$$

where $\lambda_k = \log\left(\frac{U_k}{U_k - 1}\right)$. A_k describes the number of jump times to be implanted in the interval $(\log(U_k - 1), \log U_k)$.

Step 2. Determination of position of jumps.

Let

$$B_k = \begin{cases} 0, & k = 0 \\ A_1 + \dots + A_k, & k \geq 1. \end{cases}$$

Define unordered samples

$$D_j^{(k)} = W_{B_{k-1}+j}(2) \log(U_k - 1) + (1 - W_{B_{k-1}+j}(2)) \log U_k \quad (2.4)$$

for $1 \leq j \leq A_k$, $k \geq 1$. Let

$$\log \tau_{B_{k-1}+j} = D_{(j)}^{(k)}, \quad 1 \leq j \leq A_k, \quad k \geq 1 \quad (2.5)$$

(ordered samples).

Step 3. Completion of the sequence.

Extend the sequence $\{\log \tau_n; n \geq 1\}$ to the whole time interval $(-\infty, \infty)$.

This can be done in different ways; one possibility is to construct the jump times $\{\log \tau_n; n \leq 0\}$ in 'reverse' time such that $\log \tau_0 < 0$.

This procedure requires at most another countable set of independent r.v.'s, independent from the previous ones, which again exist on the same probability space, eventually after enlarging.

3. Extremal jumps and record times

From the foregoing it is clear that the sequences $\{\log \tau_n; n \geq 1\}$ and $\{\log U_n; n \geq 1\}$ are closely related since $\log \tau_n$ only takes values in the set $\bigcup_{k=1}^{\infty} (\log(U_k - 1), \log U_k)$; particularly, there exists some a.s. finite r.v. S such that

$$\log \tau_{n+S} \in (\log(U_n - 1), \log U_n) \text{ a.s.} \quad (3.1)$$

for sufficiently large n (cf. [7],[9]). From here it follows that

$$\begin{aligned} \log U_n &= \log \tau_{n+S} + o(\exp\{-n + nH(\frac{1}{n})\}) \\ &= \log \tau_n + O(\log n) \text{ a.s. } (n \rightarrow \infty) \end{aligned} \quad (3.2)$$

where $tH(\frac{1}{t})$, $t > 0$ belongs to the upper class of a Wiener process (see [6]) since $\log U_n - \log(U_n - 1) \sim \frac{1}{U_n}$ with $n \rightarrow \infty$.

Relation (3.2) does not provide the best possible strong approximation of $\{\log U_n\}$ by a homogeneous Poisson process with unit rate. In fact, in [6] it was proved that there exists a Poisson process $\{\tau_n; n \geq 1\}$ with unit rate, defined on the same probability space as the original sequence, and a r.v. $Z \geq 0$ which is asymptotically independent of this process such that

$$\log U_n = T_n + Z + o(\exp\{-n + nH(\frac{1}{n})\}) \text{ a.s. } (n \rightarrow \infty) \quad (3.3)$$

which gives an a.s. $O(1)$ rate result. It was also shown that Z can be represented as

$$Z = \sum_{k=1}^{\infty} \log(1 + \frac{W_k}{U_k - 1}) \quad (3.4)$$

where $\{W_k\}$ is an i.i.d. sequence of $U[0,1]$ -distributed r.v.'s independent of

$\{U_k\}$, and that $E(Z) = 1 - C$ (cf. also [4]). Although $\{T_n\}$ and $\{\log \tau_n\}$ are not directly comparable, there is however a conditional relationship between Z and S , given the σ -field A generated by the record times.

Theorem 1. We have

$$E(S|A) = E(Z|A) \quad \text{a.s.}, \quad (3.5)$$

hence

$$E(S) = E(Z) = 1 - C.$$

Proof. According to what has been said in Section 2, we have

$$E(A_k|U_k) = U_k \log \left(\frac{U_k}{U_k - 1} \right) \quad \text{a.s.}$$

which is the (conditional) mean of the truncated Poisson distribution $Q(\lambda_k)$ with $\lambda_k = \log \left(\frac{U_k}{U_k - 1} \right)$. But a little analysis shows that also

$$U_k \log \left(\frac{U_k}{U_k - 1} \right) = E(\log(1 + \frac{W_k}{U_k - 1}) | U_k) + 1 \quad \text{a.s.} \quad (3.6)$$

The result now follows by the observation that $S = \sum_{k=1}^{\infty} (A_k - 1)$; hence

$$E(S|A) = \sum_{k=1}^{\infty} E(\log(1 + \frac{W_k}{U_k - 1}) | A) = E(Z|A) \quad \text{a.s.} \quad (3.7)$$

In the light of (3.1), S is the surplus number of points in $\{\tau_n; n \geq 1\}$ compared with $\{U_n; n \geq 1\}$.

It should be pointed out that since $\{\log \tau_n\}$ has i.i.d. increments following an exponential distribution with unit mean, the limiting distribution of $\log U_n - \log \tau_n$ is that of $Z^* = \sum_{k=1}^S Y_k$, where $\{Y_k\}$ are i.i.d. exponential random

variables with unit mean, independent of S , such that again $E(Z^*) = 1 - C$. However, Z^* and Z are not identical in distribution since $P(S = 0) \geq C > 0$; hence Z^* has an atom at zero, while Z has no atoms.

Since $A = \sum_{k=1}^{\infty} (A_k - 1)$ and $P(A_k \geq 2)$ is "small" one might expect that Z should be close to a Poissonian r.v. This can be precised in the following way.

Theorem 2. Let $P(\cdot)$ denote the Poisson distribution with mean $\mu > 0$. Then with $\mu = 1 - C = .422$, S is approximately $P(\mu)$ -distributed, with

$$\sup_{M \subseteq \mathbb{Z}^+} |P(S \in M) - P(\mu)(M)| \leq .12.$$

Proof. According to Theorem 3 (Appendix), if d denotes the total variation distance and P^X the distribution of a r.v. X , we have, conditionally on A , with $\lambda_k = \log(\frac{U_k}{U_k - 1})$,

$$d(P^{A_{k-1}}(\cdot | A), P(\frac{1}{2}\lambda_k)) \leq .06\lambda_k^3 \quad (3.8)$$

since $\lambda_k \leq \log(1 + \frac{1}{k}) \leq \log 2$ for all $k \geq 1$.

A little calculus shows that for any $y \geq 2$, we have

$$|y \log(\frac{y}{y-1}) - 1 - \frac{1}{2} \log(\frac{y}{y-1})| \leq \frac{1}{12} \log^2(\frac{y}{y-1}) \quad (3.9)$$

hence (cf. [11])

$$d(P(\frac{1}{2}\lambda_k), P(U_k \lambda_k - 1)) \leq \frac{1}{12} \lambda_k^2 \quad (3.10)$$

and

$$d(P^S(\cdot|A), P(E(Z|A))) \leq \frac{1}{12} \sum_{k=1}^{\infty} \lambda_k^2 + .06 \sum_{k=1}^{\infty} \lambda_k^3 \quad (3.11)$$

by (3.6) and (3.7). From here it follows that, if $E[P(E(Z|A))]$ denotes the corresponding mixed Poisson distribution, that

$$d(P^S, E[P(E(Z|A))]) \leq \frac{1}{12} \sum_{k=1}^{\infty} E(\lambda_k^2) + .06 \sum_{k=1}^{\infty} E(\lambda_k^3) \quad (3.12)$$

and hence by [5]

$$d(P^S, P(E(Z))) \leq \frac{1}{12} \sum_{k=1}^{\infty} E(\lambda_k^2) + .06 \sum_{k=1}^{\infty} E(\lambda_k^3) + \text{Var}(E(Z|A)) \quad (3.13)$$

where $E(Z) = \mu = 1 - C$. Now, for $y \geq 2$,

$$.08136 \log^2 \left(\frac{y}{y-1} \right) \leq |1 - (y-1) \log^2 \left(\frac{y}{y-1} \right) - (y-1)^2 \log^2 \left(\frac{y}{y-1} \right)| \leq \frac{1}{12} \log^2 \left(\frac{y}{y-1} \right),$$

hence

$$\begin{aligned} \text{Var}(E(Z|A)) &= \text{Var}(Z) - E(\text{Var}(Z|A)) \\ &= \text{Var}(Z) - \sum_{k=1}^{\infty} \{1 - E[(U_k - 1)\lambda_k^2] - E[(U_k - 1)^2 \lambda_k^2]\} \quad (3.14) \\ &\leq \text{Var}(Z) - .08136 \sum_{k=1}^{\infty} E(\lambda_k^2), \end{aligned}$$

hence

$$d(P^S, P(\mu)) \leq .002 \sum_{k=1}^{\infty} E(\lambda_k^2) + .06 \sum_{k=1}^{\infty} E(\lambda_k^3) + \text{Var}(Z).$$

But $\lambda_k \leq \log(1 + \frac{1}{k})$, and $\text{Var}(Z) \leq .09$, hence the result follows by some numerical computations.

Clearly, by the a.s. finiteness of S , we can now also construct the extremal process $E(t)$ interpolating the given partial maxima sequence, at least for $t \geq \tau_{S+T}$ where $T = \inf\{k | \tau_{S+k} \in (U_k - 1, U_k)\}$ is also a.s. finite. We only have to define

$$E(t) = \sum_{k=T}^{\infty} X_{U_k} I_{[\tau_{S+k}, \tau_{S+k+1})}(t), \quad t \geq \tau_{S+T}. \quad (3.15)$$

Then $E(n) = X_{(n)}$ for all $n \geq \tau_{S+T}$. If, for example, $F = \Lambda$, then

$X_{([nt])} - \log n \approx E(nt) - \log n$ for $t \geq \tau_{S+T}/n$ which now is extremal- Λ on the larger interval $(\tau_{S+T}/n, \infty)$, similarly for general F . This provides another strong approximation approach for the limiting extremal processes as worked out in [1],[2].

4. Appendix

Theorem 3. Let the r.v. X have a truncated Poisson distribution $Q(\lambda)$ with $\lambda > 0$, and let Y have a Poisson distribution $P(v)$ with $v = \frac{\lambda}{2}$. Then, if $\lambda \leq \lambda_0 = 2.702$ (the root of $\sinh(\frac{\lambda_0}{2}) = \frac{2}{3}\lambda_0$), we have

$$\begin{aligned} d(P^{X-1}, P^Y) &= \sup_{M \subseteq \mathbb{Z}^+} |P(X-1 \in M) - P(Y \in M)| \\ &= \frac{1}{e^{\frac{\lambda}{2}} - 1} \{2 \sinh(\frac{\lambda}{2}) - \lambda\} \{1 + \frac{\lambda}{2}\} \end{aligned} \quad (4.1)$$

which in turn can be estimated by

$$\frac{\lambda^3}{48} \frac{(2+\lambda) \cosh(\frac{\lambda}{6})}{e^{\frac{\lambda}{2}} - 1}. \quad (4.2)$$

Proof. The sup in the total variation distance is attained for the set $M = \{k \in \mathbb{Z}^+ \mid P(X=k+1) \leq P(Y=k)\}$, which is $M = \{0,1\}$ in the case $\lambda \leq \lambda_0$. This follows from the fact that

$$P(Y=k) = e^{-v} \frac{v^k}{k!} \geq \frac{1}{e^{\frac{\lambda}{2}} - 1} \frac{\lambda^{k+1}}{(k+1)!}$$

if and only if

$$\sinh(\frac{\lambda}{2}) \geq \frac{2^{k-1}}{k+1} \lambda, \quad k \geq 0.$$

By our assumption, this is only true for $k=0,1$, which gives (4.1). (4.2) follows from the Taylor expansion for \sinh .

REFERENCES

- [1] Deheuvels, P. (1981): The strong approximation of extremal processes. *Z. Wahrscheinlichkeitsth. verw. Geb.* 58, 1 - 6.
- [2] Deheuvels, P. (1982): Strong approximation in extreme value theory and applications. *Coll. Math. Soc. János Bolyai* 36, Limit Theorems in Probability and Statistics, Veszprém, Hungary. North Holland, Amsterdam, 369 - 403.
- [3] Deheuvels, P. (1983): The complete characterization of the upper and lower class of the record and inter-record times of an i.i.d. sequence. *Z. Wahrscheinlichkeitsth. verw. Geb.* 62, 1 - 6.
- [4] Pfeifer, D. (1984): A note on moments of certain record statistics. *Z. Wahrscheinlichkeitsth. verw. Geb.* 66, 293 - 296.
- [5] Pfeifer, D. (1985): On the distance between mixed Poisson and Poisson distributions. Center for Stochastic Processes, UNC at Chapel Hill, Tech. Rep. # 115.
- [6] Pfeifer, D. (1985): On a joint strong approximation theorem for record and inter-record times. Center for Stochastic Processes, UNC at Chapel Hill, Tech. Rep. # 120.
- [7] Resnick, S.I. (1973): Extremal processes and record value times. *J. Appl. Prob.* 10, 863 - 868.
- [8] Resnick, S.I. (1974): Inverses of extremal processes. *Adv. Appl. Prob.* 6, 392 - 400.
- [9] Resnick, S.I. (1975): Weak convergence to extremal processes. *Ann. Prob.* 3, 951 - 960.
- [10] Resnick, S.I. and Rubinovitch, M. (1973): The structure of extremal processes. *Adv. Appl. Prob.* 5, 287 - 307.
- [11] Serfling, R.J. (1978): Some elementary results on Poisson approximation in a sequence of Bernoulli trials. *SIAM Review* 20, 567 - 579.
- [12] Westcott, M. (1977): A note on record times. *J. Appl. Prob.* 14, 637 - 639.
- [13] Williams, D. (1973): On Rényi's 'record' problem and Engel's series. *Bull. Lond. Math. Soc.* 5, 235 - 237.

END

DTIC

8-86